

Прогностическое моделирование скорости движения атмосферы на основе стохастических регрессоров

Громковский А. А. e-mail: aag68@bk.ru
Костылева Л. Н. e-mail: kostyleva12@yandex.ru
Мозиков Б. В.

ФГКВООУ ВО ВУНЦ ВВС «Военно-воздушная академия им. проф. Н. Е. Жуковского и Ю. А. Гагарина», г. Воронеж, Россия

***Аннотация.** На основе пространственно-распределенных данных метеорологических наблюдений скорости ветра построена прогностическая регрессионная модель скорости движения атмосферы. Анализ построенной модели выполнен на основе проверки статистических гипотез о параметрах модели. Алгоритм проверки статистических гипотез о параметрах модели реализован на основе предположения о стохастической природе независимых переменных-регрессоров прогностической модели. Предлагаемый подход упрощает выявление и оценку причинно-следственной связи зависимой и независимых переменных при осуществлении прогнозирования. На основе данного подхода можно достаточно просто строить интервальные оценки параметров прогностической модели.*

***Ключевые слова:** Моделирование, прогнозирование, асимптотический подход в регрессионном анализе, стохастические предикторы прогностической модели, оценивание причинно-следственной связи переменных прогностической модели, анализ модели.*

Введение.

Краткосрочный прогноз скорости движения атмосферы можно построить с использованием подхода прогностического моделирования [1]. Прогностическая модель скорости ветра может быть построена с применением методов регрессионного анализа на основе данных пространственно-распределенных выборочных совокупностей метеорологических наблюдений для данного участка атмосферы [1, 2]. При построении прогностической модели важно учитывать природу независимых переменных – регрессоров. Это имеет принципиальное значение при проведении анализа модели и оценки значимости причинно-следственной связи прогнозируемой переменной и переменных-регрессоров.

1. Отбор прогностических регрессоров для построения модели.

Для построения прогностической регрессионной модели скорости движения атмосферы была сформирована архивная выборочная совокупность за период с 2014 по 2018 год по метеостанции города Пермь в количестве 10000 случаев. В качестве факторов изменения скорости движения атмосферы F_f (м/с) рассматривались температура воздуха T (К), температура точки росы T_d (К), влажность воздуха U (%), атмосферное давление (Па), атмосферное давление, приведенное к уровню моря P_0 (Па), барометрическая тенденция P_a (Па), направление ветра VV (град.) и количество баллов облачности N . После выборочная совокупность была разделена на обучающую и контрольную в отношении 2:1, обучающая выборка составила 6667 случаев, а контрольная 3333 случая.

Отбор предикторов для прогностической регрессионной модели скорости движения атмосферы осуществлялся посредством реализации процедуры последовательного присоединения (просеивания) на основе анализа матрицы коэффициентов парной корреляции [1]. Корреляционная матрица для исследуемых величин представлена в приведенной ниже таблице.

Таблица

Корреляционная матрица регрессоров скорости движения атмосферы

	F_f	T	P_0	P	P_a	U	N	VV	T_d
F_f	1								
T	0,01	1							
P_0	-0,52	-0,25	1						
P	-0,22	-0,33	0,99	1					
P_a	-0,48	-0,06	0,05	0,06	1				
U	-0,17	-0,33	-0,15	-0,12	0,10	1			
N	0,44	-0,03	-0,24	-0,23	-0,07	0,24	1		
VV	-0,05	0,31	0,20	0,17	0,03	-0,44	-0,33	1	
T_d	-0,05	0,93	-0,31	-0,38	-0,02	0,01	0,04	0,18	1

Проведенная оценка корреляции показывает, что величины количество баллов облачности N , атмосферное давление, приведенное к уровню моря P_0 и барометрическая тенденция P_a демонстрируют умеренную тесноту взаимосвязи со скоростью ветра F_f . Остальные регрессоры показывают слабую и очень слабую связь, следовательно их нецелесообразно включать в модель.

2. Оценивание параметров прогностических моделей.

Дальнейшая реализация метода последовательного присоединения заключалась в переборе моделей с различным количеством отобранных регрессоров. На основе обучающей выборочной совокупности были построены прогностические модели для скорости ветра Ff с одним, двумя и тремя регрессорами. Параметры прогностических моделей оценивались методом наименьших квадратов [2, 3].

На первом этапе последовательного присоединения было построено три прогностических модели с одним регрессором:

$$Ff = 29,67 - 0,03 P_0 \quad (1)$$

$$Ff = 2,38 - 0,19 P_a \quad (2)$$

$$Ff = 1,64 + 0,09 N \quad (3)$$

На втором этапе последовательного присоединения было построено три двухфакторной модели:

$$Ff = 1,68 - 0,16 P_a + 0,09 N \quad (4)$$

$$Ff = 28,82 - 0,03 P_a - 0,17 P_0 \quad (5)$$

$$Ff = 23,13 - 0,03 P_0 + 0,08 N \quad (6)$$

На завершающем этапе последовательного присоединения была построена трехфакторная прогностическая модель:

$$Ff = 22,59 - 0,03 P_0 - 0,17 P_a + 0,08 N \quad (7)$$

На основе данных контрольной выборочной совокупности были определены основные показатели успешности прогностических моделей: средняя абсолютная ошибка (CAO), среднеквадратичное отклонение (СКО), коэффициент корреляции (R_{xy}) [1, 4].

На первом этапе были рассчитаны оценки качества для прогностических моделей парной регрессии:

Модель $Ff = 29,67 - 0,03 P_0$, CAO=4,89; СКО=6,95; R_{xy} =0,21.

Модель $Ff = 2,38 - 0,19 P_a$, CAO=2,76; СКО=4,12; $R_{xy}=0,18$.

Модель $Ff = 1,64 + 0,09 N$, CAO=3,01; СКО=5,05; $R_{xy}=0,21$.

Для данных прогностических моделей ошибка прогноза составляет от 3 до 7 м/с. Низкий коэффициент корреляции прогнозирования скорости движения атмосферы является существенной причиной отвержения модели в целях прогнозирования. Модели парной регрессии нецелесообразно использовать в метеорологической практике.

Двухфакторные прогностические модели оценивались по тем же критериям.

Модель $Ff = 1,68 - 0,16 P_a + 0,09 N$, CAO=2,54; СКО=3,98; $R_{xy}=0,28$.

Модель $Ff = 28,82 - 0,03 P_a - 0,17 P_0$, CAO=2,36; СКО=3,75; $R_{xy}=0,32$.

Модель $Ff = 23,13 - 0,03 P_0 + 0,08 N$, CAO=2,98; СКО=4,04; $R_{xy}=0,26$.

Для двухфакторных прогностических моделей ошибка прогноза составляет от 2,5 до 4 м/с, коэффициент корреляции низкий. Двухфакторные прогностические модели скорости движения атмосферы нецелесообразно использовать для прогнозирования скорости ветра в метеорологическом подразделении.

Для трехфакторной прогностической регрессионной модели $Ff = 22,59 - 0,03 P_0 - 0,17 P_a + 0,08 N$ средняя абсолютная ошибка составила 1,02, среднеквадратичное отклонение составило 2,01, коэффициент корреляции 0,52. Ошибка трехфакторной прогностической модели находится в пределах 1 – 2 м/с. При использовании данной прогностической регрессионной модели для прогнозирования скорости движения атмосферы предполагается реализация оправдываемых в большинстве случаев прогнозов.

3. Оценка значимости прогностической регрессионной модели скорости ветра в целом.

Коэффициент детерминации полученной прогностической регрессионной модели составил 0,39. Полученное прогностическое уравнение на 39% объясняет влияние включенных в него регрессоров скорость движения атмосферы. Оставшиеся 61% влияния можно объяснить факторами, которые не учитывались при прогнозировании.

Статистическая значимость прогностической регрессионной модели скорости ветра проверялась на основе следующих гипотез [3]:

$H_0: R^2 = 0$ – построенное уравнение в целом незначимо;

$H_1: R^2 \neq 0$ – построенное уравнение в целом значимо.

Для проверки статистической значимости построенной модели регрессии в целом использовалась статистика одностороннего критерия

Фишера $F_{табл}(\alpha; m; n - m - 1)$, где α – уровень значимости; m – количество параметров модели; n – количество наблюдений. Табличное значение статистики критерия Фишера для уровня значимости в 1% составило $F_{табл}(0,01; 3; 6663) = 3,78$.

Расчетное значение статистики критерия Фишера как отношение объясненной и остаточной дисперсии зависимой переменной составило $F_{расч} = 237,9$.

Так как $F_{расч} > F_{табл}$ на уровне значимости (1%) принимается гипотеза H_1 . Построенная прогностическая модель в целом значима, получено надежное математическое описание прогноза скорости движения атмосферы.

4. Оценка значимости параметров прогностической регрессионной модели скорости ветра.

Для проверки статистической значимости параметров прогностической регрессионной модели скорости движения атмосферы выдвигалась пара гипотез для каждого из параметров. Основная – о незначимом отличии от нуля полученной оценки данного параметра, альтернативная – о значимом отличии от нуля полученной оценки параметра [4].

$H_0: a_j = 0$, параметр незначим;

$H_1: a_j \neq 0$ параметр значим.

Для проверки выдвинутых гипотез использовалась статистика критерия Стьюдента.

Табличное значение статистики критерия Стьюдента $t_{табл}(0,01; 6667) = 2,57$.

Расчетное значение статистики критерия Стьюдента параметра a_0 :

$$t_{a_0}^{расч} = \frac{a_0}{m_{a_0}} 15,23; \quad \left| t_{a_0}^{расч} \right| > t_{табл}.$$

Для параметра a_0 при заданном уровне значимости принимается гипотеза H_1 , коэффициент несет значимую информацию в модели.

Расчетное значение статистики критерия Стьюдента параметра a_1 :

$$t_{a_1}^{расч} = \frac{a_1}{m_{a_1}} = -14,11; \quad \left| t_{a_1}^{расч} \right| > t_{табл}.$$

Для параметра a_1 при заданном уровне значимости принимается гипотеза H_1 , коэффициент несет значимую информацию в модели.

Расчетное значение статистики критерия Стьюдента параметра a_2 :

$$t_{a_2}^{расч} = \frac{a_2}{m_{a_2}} = 15,86 ; \left| t_{a_2}^{расч} \right| > t_{табл} . \text{ Для параметра } a_2 \text{ при заданном}$$

уровне значимости принимается гипотеза H_1 , коэффициент несет значимую информацию в модели.

Расчетное значение статистики критерия Стьюдента параметра a_3 :

$$t_{a_3}^{расч} = \frac{a_3}{m_{a_3}} = -8,73 ; \left| t_{a_3}^{расч} \right| > t_{табл} . \text{ Для параметра } a_3 \text{ при заданном}$$

уровне значимости принимается гипотеза H_1 , коэффициент несет значимую информацию в модели.

Все параметры построенной прогностической регрессионной модели скорости движения атмосферы значимы. Получены надежные статистические оценки параметров, содержащие объективную информацию о влиянии регрессоров на прогнозируемую переменную.

5. Анализ прогностической модели скорости движения атмосферы в предположении стохастичности предикторов.

Классическая модель множественной регрессии может быть построена в предположении стохастичности переменных-регрессоров. В этом случае предпосылки применения данной модели ослабляются и работает асимптотический подход [2, 3, 5]. Стохастичность регрессоров корректно отражает физическую сущность метеорологических величин, использованных для построения прогностической модели скорости движения атмосферы. Поэтому целесообразно использовать тестирование значимости параметров построенной прогностической регрессионной модели скорости ветра в предположении асимптотического подхода.

Основная гипотеза $H_0: a_j = 0$, альтернативная гипотеза $H_1: a_j \neq 0$.

Алгоритм проверки выдвинутой гипотезы будем следующим.

1. Расчетное значение тестовой статистики определяется по формуле $u = \frac{a_j}{m_{aj}}$.

2. Определяется P – значение стандартного нормального распределения . $P = 2\Phi\left(-\left|\frac{a_j}{m_{aj}}\right|\right)$

3. Определяется уровень значимости α .

4. Если P – значение меньше уровня значимости $2\Phi\left(-\left|\frac{a_j}{m_{aj}}\right|\right) < \alpha$

то следует отвергнуть основную гипотезу H_0 и сделать вывод в пользу альтернативной гипотезы,

Соответственно, на уровне значимости 1% можно получить следующее условие отвержения выдвинутой гипотезы:

$$2\Phi\left(-\left|\frac{a_j}{m_{aj}}\right|\right) < 0,01. \text{ Это условие эквивалентно равенству } \left|\frac{a_j}{m_{aj}}\right| > 2,58.$$

С учетом определенных ранее отношений оценок параметров моделей к стандартным ошибкам их оценивания видно, что все параметры прогностической модели скорости движения атмосферы в предположении стохастичности статистически значимы.

6. Интервальное оценивание параметров прогностической модели движения атмосферы со стохастическими регрессорами.

Для построения интервальных оценок параметров прогностической модели скорости движения атмосферы были определены критические значения тестовой статистики при уровне значимости 1%, которое составляет 2,58 и при уровне значимости 5%, которое составляет 1,96.

На первом этапе оценивания были рассчитаны интервальные оценки всех параметров при уровне значимости 1%:

Для параметра α_0 , точечная оценка равна 22,59 стандартная ошибка составляет 1,08. Данный параметр находится в интервале от 18,76 до 26,42.

Для параметра α_1 , точечная оценка равна -0,03 стандартное отклонение составляет -0,0019. Данный параметр находится в интервале от -0,32 до -0,27.

Для параметра α_2 , точечная оценка равна -0,15 стандартное отклонение составляет -0,017. Данный параметр находится в интервале от -0,2 до -0,11.

Для параметра α_3 , точечная оценка равна 0,08 стандартное отклонение составляет -0,0048. Данный параметр находится в интервале от 0,08 до 0,63.

На втором этапе оценивания были рассчитаны интервальные оценки всех параметров при уровне значимости 5%:

Для параметра α_0 , точечная оценка равна 22,59 стандартное отклонение составляет 1,08. Данный параметр находится в интервале от 19,68 до 25,5.

Для параметра α_1 , точечная оценка равна -0,03 стандартное отклонение составляет -0,0019. Данный параметр находится в интервале от -0,031 до -0,024.

Для параметра α_2 , точечная оценка равна -0,15 стандартное отклонение составляет -0,017. Данный параметр находится в интервале от -0,18 до -0,12.

Для параметра α_3 , точечная оценка равна 0,08 стандартное отклонение составляет -0,0048. Данный параметр находится в интервале от 0,07 до 0,09.

Заключение.

Построение прогностической модели метеорологической величины в предположении стохастичности регрессоров на основе использования асимптотического подхода позволяет упростить процедуру оценки значимости параметров модели.

На основе формул проверки значимости параметров можно достаточно просто определить интервальные оценки параметров модели. Интервальное оценивание параметров прогностической регрессионной модели является более корректным способом вычисления значений показателей модели на основе экспериментальных данных, по сравнению с точечным оцениванием. Асимптотическое предположение о характере предикторов создаваемой модели позволяет упростить вычисления интервальных оценок по сравнению с подходом классической линейной модели множественной регрессии.

Список литературы

1. Блейк, Д. Физические основы динамики атмосферы и метеорологии / Д. Блейк, Р. Робсон. – Долгопрудный.: ИД «Интеллект», 2016. – 160 с.
2. Горяинова Е.Р. Прикладные методы анализа статистических данных / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. – М.: Изд. дом Высшей школы экономики, 2012. – 310 с.
3. Дрейпер, Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М.: ООО «И.Д. Вильямс», 2017. – 912 с.
4. Кун, М. Предиктивное моделирование на практике / М. Кун, К. Джонсон. – СПб.: Питер, 2019. – 640 с.
5. Матвеев, М. Г. Разработка и исследование статистических моделей нестационарного многомерного временного ряда атмосферных температур в условиях неоднородности / М. Г. Матвеев, Е. А. Сирота // Информационные технологии. – 2014. – № 12. – С. 20-24.